***Xenopus laevis* 18S ribosomal RNA: experimental determination of secondary structural elements, and locations of methyl groups in the secondary structure model**

Johannes Atmadja and Richard Brimacombe

Max-Planck-Institut für Molekulare Genetik, Abteilung Wittmann, D-1000, Berlin-Dahlem, FRG, and

B. Edward H. Maden*

Department of Biochemistry, University of Glasgow, Glasgow G12 8QQ, UK

Received 20 January 1984; Accepted 28 February 1984

SUMMARY

18S ribosomal RNA from *X. laevis* was subjected to partial digestion with ribonucleases A or T₁ under a variety of conditions, and base-paired fragments were isolated. Sequence analysis of the fragments enabled five base-paired secondary structural elements of the 18S RNA to be established. Four of these elements (covering bases 221-256, 713-757, 1494-1555 and 1669-1779) confirm our previous secondary structure predictions, whereas the fifth (comprising bases 1103-1125) represents a phylogenetically conserved "switch" structure, which can also form in prokaryotic 16S RNA. The results are incorporated into a refined model of the 18S RNA secondary structure, which also includes the locations of the many methyl groups in *X. laevis* 18S RNA. In general the methyl groups occur in non-helical regions, at hairpin loop ends, or at helix boundaries and imperfections. One large cluster of 2'-O-methyl groups occurs in a region of complicated secondary structure in the 5'-one third of the molecule.

INTRODUCTION

A number of different experimental approaches have been applied to the problem of secondary structure analysis in large ribosomal RNA molecules (see ref. 1 for review). Of these, the most direct is the "base-paired fragment" approach (2-4), in which partial nuclease digests of the RNA are separated by two-dimensional gel electrophoresis using non-denaturing conditions in the first dimension and denaturing conditions in the second. The DNA corresponding to the 18S ribosomal RNA from *Xenopus laevis* has recently been sequenced (5-7), and in this paper we describe the analysis of a number of base-paired fragments isolated from the *X.laevis* 18S RNA.

The study was undertaken for three reasons. Firstly, although phylogenetic extrapolations (cf.8) from the corresponding *E. coli* structures have been made by several authors in order to

derive secondary structure models of eukaryotic 18S (9-12) or 26S (13) RNA, there is so far only very limited experimental evidence available in support of the eukaryotic structures (e.g. ref. 12). Secondly, we hoped that direct structural information gained from the 18S RNA could be extrapolated back into the 16S RNA, thus leading to a greater insight into the E. coli structure. Thirdly, eukaryotic ribosomal RNA contains numerous methyl groups (5,7,14-16) and other modified nucleotides (17-19), whose locations in relation to secondary structure models are for the most part undefined. It is likely that knowledge of the locations of the modified nucleotides will add to our understanding of features of secondary and tertiary structure, and vice versa.

The nuclease digestion data obtained define five distinct secondary structural elements in the X. laevis 18S RNA; four of these confirm helices that we had previously proposed (9), whereas the fifth is a new interaction. The results are incorporated into a refined secondary structure model of the entire 18S RNA molecule, and are discussed in relation both to the latest E. coli 16S RNA structure (20), and to the locations of the methyl groups in X. laevis 18S RNA.

MATERIALS AND METHODS

Isolation of 18S ribosomal RNA from X. laevis cells

18S ribosomal RNA was isolated from ³²P-labelled or unlabelled cultured X. laevis kidney cells using minor modifications of previously described procedures (15,21). In outline, cells were harvested by trypsinization, extracted with phenol at room temperature, and material from the aqueous phase was subjected to sucrose gradient centrifugation as described (15), yielding complete separation of 18S from 28S RNA. Labelled and unlabelled RNA were mixed to give a specific activity of the order of 1.5×10^6 ³²P-counts/min per A₂₆₀ unit.

Partial digestion of 18S RNA and separation of RNA fragments

Aliquots of 18S RNA containing 1 A₂₆₀ unit were digested in 20 µl of a buffer containing 10 mM Tris-HCl pH 7.8, 10 mM NaCl, 0.1% sodium dodecyl sulphate, and either 0.3 mM EDTA or 0.3 mM magnesium acetate. The digestions were made with either (a) ribonuclease T₁ (3-300 enzyme units per aliquot) or (b) ribo-

nuclease A (10 - 200 ng per aliquot) or (c) mixtures of both enzymes (10 - 30 units of ribonuclease T₁, and 100 - 300 ng of ribonuclease A per aliquot). Ribonuclease T₁ digestions were carried out for 30 min at either room temperature or 37°, ribonuclease A digestions for 30 min at 37°, and the mixed digestions for 60 min at 4°. In each case the reaction was stopped by incubation with 10 µg proteinase K (Merck, Darmstadt) for 15 min at room temperature. The digested samples were loaded onto the two-dimensional gel system, the first dimension being a 5 - 20% polyacrylamide gradient gel containing dodecyl sulphate and the second a 25% polyacrylamide gel in the buffer system of Maxam and Gilbert (22), as previously described (2,3). After electrophoresis the RNA fragments were visualized by autoradiography, extracted from the gels, and stored at -80° until required (23).

Since the specific activity of the in vivo ³²P-labelled RNA could not be made high enough to allow direct "fingerprinting" of the isolated RNA fragments, the ³²P-radioactivity was deliberately kept to a relatively low level (see above), sufficient only to enable the fragments to be located on the gel by autoradiography. The subsequent sequence analysis was made by post-labelling of the individual isolated RNA fragments.

5'-end labelling and sequencing of RNA fragments

Individual RNA fragments were labelled with ³²P-ATP and polynucleotide kinase exactly as described (23). The labelled fragments were purified by electrophoresis on 25% gels in the Maxam and Gilbert system (22), localized by autoradiography and extracted from the gels as above. Sequencing was carried out by the enzymatic method of Donis-Keller et al. (24, and cf. ref. 23). For the sequencing reactions, 10 µl aliquots of 5'-labelled fragment containing 10 µg of unlabelled carrier RNA were digested with the following enzymes for 30 min at 37° :- (a) ribonuclease A, 2 or 6 ng per aliquot, (b) ribonuclease T₁, 1 or 2 enzyme units per aliquot, (c) ribonuclease Phy M, 4 units per aliquot, (d) ribonuclease B. cereus, 2 units per aliquot, and (e) ribonuclease U₂, 0.03 or 0.06 units per aliquot. Digestions (a) to (d) were made in 10 mM triethanolamine-HCl pH 7.8, and digestion (e) in 3 mM triethanolamine-acetate pH 4.5. Ribonuclease A and T₁ were obtained from the Sigma Chemical Co., and the remaining

nucleases from PL Biochemicals. The digested samples were loaded onto 25% polyacrylamide gels (22) as above, together with an aliquot which had been digested for 30 min at 90° in 50 mM NaHCO₃/Na₂CO₃ buffer pH 9.0 containing 1 mM EDTA. A further aliquot (untreated) was also included as a control. The gels were run until a xylene cyanol blue marker had travelled approximately 15 cm, and were then subjected to autoradiography. In cases where the sequence could not be followed to the extreme 3'-end of the fragment as a result of structure forming in the 3'-region, the position of the 3'-end was estimated (to within ± 2 nucleotides) from the mobility of the unhydrolysed fragment in the sequencing gel (cf. ref. 23).

RESULTS AND DISCUSSION

Identification of base-paired fragments

Some examples of the two-dimensional gel separations obtained from partial digests of the 18S RNA (see Materials and Methods) are illustrated in Fig. 1. Fig. 1A shows a particularly clear pattern, with the expected diagonal of unstructured fragments or uncut hairpin loops and three sets of fragment pairs lying below the diagonal (cf. refs. 2,3). Of these three sets, fragments 2a and 2b represent an unmistakable "classical" pair of fragments of approximately equal length, whereas fragment pair 1 is composed of one long fragment (1a) together with a shorter piece (1b). Fragment pair 5 appears as three spots, and the subsequent analysis showed that the central strong spot was a homogeneous fragment (5a), whereas the two weaker flanking spots correspond to the complementary sequence (5b) in two versions differing in length by two to three nucleotides.

The gel pattern of Fig. 1A is very similar to the type of result usually obtained with *E. coli* RNA (3,4), but in this series of experiments (perhaps as a result of the rather high G-C content of parts of the *X. laevis* RNA) the gel patterns tended to be more complicated, as exemplified by the autoradiograms of Figs. 1B and 1C. In such cases complex groups of fragments lying below the diagonal were found, but nevertheless a number of obvious fragment pairs could usually also be discerned (see the numbered spots in Figs. 1B, 1C).

Figure 1: Two-dimensional gel separation of RNA fragments. Direction of electrophoresis is from right to left (first dimension) and top to bottom (second dimension). The dashed lines represent the "diagonal" (cf. refs. 2,3), and the numbers assigned to the spots correspond to those of Table 1. Essentially identical fragments, prolonged by one to three bases, are indicated by the corresponding numbers in parentheses. A: Fragments derived from a ribonuclease T₁ digest (30 units enzyme, 0.3 mM magnesium, room temperature (cf. Materials and Methods)). B: Fragments derived from a mixed enzyme digest (20 units ribonuclease T₁, 200 ng ribonuclease A, 0.3 mM EDTA). C: As for B, but with 30 units ribonuclease T₁ and 100 ng ribonuclease A. (Fragments Xa and Xb were not analysed, see text).

A large number of fragment pairs from the various two-dimensional gels were subjected to 5'-end-labelling and sequence analysis as described in Materials and Methods. The yield of the 5'-labelling reaction with RNA fragments is known to be very variable and sequence specific (25), and in a few cases the uptake of label in the individual fragments was reproducibly very poor. For this reason one or both components of some fragment pairs could not be analysed (e.g. fragment pair "X" in Fig. 1C). In general the labelling was however adequate, and examples of

Figure 2: Sequencing gels of fragments 1a (left) and 1b (right). The gel slots contained the following digests: 1, control (untreated); 2, ribonuclease *B. cereus*; 3, 4, ribonuclease A; 5, alkali "ladder"; 6, 7, ribonuclease U₂; 8, 9, ribonuclease T₁; 10, ribonuclease Phy M. See Materials and Methods for details. The sequences correspond to positions 215-238 and 239-249 in the *X. laevis* sequence (5), respectively. In the former case the position of the 3'-end was determined from a plot of chain length versus mobility in the gel (see Materials and Methods); in the latter the actual 3'-cutting site appears to be at position C-250 (implying a ribonuclease A contamination in the original ribonuclease T₁ digestion of the 18S RNA (cf. Fig. 1A)).

the sequencing gels obtained from one fragment pair are shown in Fig. 2. The sequences found were located in the *X. laevis* 18S RNA sequence (5). It should be noted that the latter sequence has been revised relative to that given in refs. 5-7, due to the discovery of a hitherto undetected A-residue following G-684. This has necessitated a change in the nucleotide numbering "downstream" from this point (B.E.H.Maden, manuscript in preparation, and see also legend to Fig. 4, below).

The fragment sets were found to originate from five distinct regions of the RNA, and the data are summarized in Table 1. As noted in Table 1, and as can be seen in Fig. 1, a number of fragment pairs appeared several times in slightly different form (the

Table 1: Base-paired fragments isolated from *X. laevis* 18S RNA. The fragment pairs (cf. Fig. 1) are listed according to their 5'- and 3'-components, denoted by the letters "a" and "b" for each pair, respectively. The enzyme used to generate the fragments from the 18S RNA is also given (cf. Materials and Methods). The interaction encompassing bases ca. 700 - 760 was found in many fragment pairs with slightly varying lengths; the two pairs listed (numbers 2 and 3, cf. Figs. 1A, 1C) are typical. Fragment pair 4 also appeared in slightly varying forms (plus or minus two bases, cf. Fig. 1B). In fragment pair 9 a third component (bases 1052-1083) was also observed.

5'-Component		3'-Component		Ribonuclease used
No.	Sequence region	No.	Sequence region	
1a	215 - 238	1b	239 - 249	T ₁
2a	709 - 735	2b	736 - 759	T ₁
3a	715 - 737	3b	740 - 760	T ₁ + A
4a	1097 - 1115	4b	1116 - 1133	T ₁ + A
5a	1487 - 1524	5b	1525 - 1556	T ₁
6a	1487 - 1506	6b	1525 - 1558	T ₁
7a	1506 - 1526	7b	1531 - 1548	A
8a	1668 - 1702	8b	1740 - 1777	T ₁
9a	1652 - 1695	9b	1740 - 1783	T ₁
10a	1692 - 1723	10b	1727 - 1761	A

individual fragments differing in length by a few nucleotides). In such cases Table 1 only lists typical specimens (e.g. fragment pairs 2,3). In other cases, significantly different sets of fragments from the same RNA region were found (pairs 5 - 7, and 8 - 10), and these are individually listed.

Secondary structures of the fragment pairs

The double-helical structures proposed for the fragments listed in Table 1 are illustrated in Fig. 3, where they are compared with the corresponding structural elements in rat (26) and/or *Saccharomyces cerevisiae* (27) 18S RNA. The experimental data confirm four secondary structural elements which we had previously proposed (9), namely Helices 11A, 21C, 41 and 44. The helices are numbered according to their positions in the complete 18S secondary structure model (see Fig. 4, below).

Helix 11A is progressively extended from *S. cerevisiae* to *X. laevis* to rat (Fig. 3), and the seven base pairs adjacent

Figure 3: Base-pairing proposed for the fragment pairs in *X. laevis* RNA (cf. Table 1). Helix numbers correspond to those of Fig. 4, and the helices are drawn in approximately the same orientations as in Fig. 4. The 5'- and 3'-ends of the individual fragments are indicated by their numbers (Table 1), together with small arrows denoting the enzyme cutting sites and arrowheads showing the direction in which each fragment runs. Arrows with dotted lines indicate the 5'-3' direction of the RNA chains. For Helix 11A, the corresponding structures for *S. cerevisiae* (27) and rat (26) 18S RNA are also shown (in the latter case the sequence is numbered as in the publication (26)). For Helix 21C, the corresponding loop in *S. cerevisiae* (bases 695-740) is given in ref. 9. In the remaining Helices (26A, 41, 44) the base changes in *S. cerevisiae* are superimposed on the *X. laevis* structures. Base changes in square boxes are compensating, whereas those in round boxes are mis-matches or in single-stranded regions. Dotted lines or "crossed-out" base pairs indicate modified base-pairing in *S. cerevisiae*. The sequence regions concerned in the latter are bases 1064-1101 (Helix 26A), 1456-1528 (Helix 41) and 1618-1746 (Helix 44).

to the closed end of the hairpin are clearly consistent with the fragment pair 1a, 1b in X. laevis. The five base pairs further away from the loop end are supported by a C-G to U-A base change in S. cerevisiae (at positions 223/244 in the latter sequence).

Helix 21C is consistent with the fragment pairs 2 and 3 (and a number of other very similar pairs (Table 1)), although it should be noted that under the more stringent digestion conditions nuclease cuts within the proposed double-helical region were sometimes observed (e.g. the 5'-ends of fragments 3a and 3b). The same phenomenon was also observed with E. coli RNA (3, 4). This helix is in a very variable part of the 18S sequence (with no counterpart in 16S RNA, cf. Fig. 4, below). However, a very similar although not directly superimposable loop can be drawn for the corresponding region of the S. cerevisiae RNA (9).

The data for Helix 41 are rather more complex. Fragment pair 5 provides a clear confirmation of the hairpin loop as a whole, whereas fragment pairs 6 and 7 are consistent with the principal sub-sections of the helix, respectively (Fig. 3). As drawn in Fig. 3, the base-pairing scheme is the same as that previously proposed (9) for X. laevis, but slightly different to that proposed for S. cerevisiae (9), and shows both a number of compensating base changes and some mis-matches between the two sequences.

Helix 44 is well confirmed by the fragment data. Fragment pairs 8 and 9 define the region of the helix distal to the hairpin loop end, whereas pair 10 defines the region nearest to the loop end. The helix (which is drawn with two extra base pairs, as compared with the previous version (9)) contains many compensating base changes between S. cerevisiae, rat and X. laevis. Furthermore, evidence has recently been published from a detailed analysis of enzyme cutting sites under mild conditions (28), which also supports the structure drawn for Helix 44. In the case of fragment pair 9, a third RNA component was observed (bases 1052-1083, Table 1) on one of the two-dimensional gels. A base-paired structure can be drawn between the latter bases and the 5'-region of fragment 9a (which includes the highly conserved bases 1651-1670), but this remains as yet unconfirmed and will not be discussed further here.

Figure 4: The revised secondary structure model of *X. laevis* 18S RNA. The sequence is divided into two domains ("A", facing page; "B", this page) and the corresponding domains of the *E. coli* structure (20) are shown as the small inset diagrams. The sequences are numbered from their 5'-ends, and corresponding helices have the same number in the *E. coli* and *X. laevis* structures. Extra helices in the *X. laevis* model are denoted by letters (11A, 11B, etc.). An extra A-residue following G-684 in the original *X. laevis* sequence (5) has been observed (see text); this change is incorporated into the structure and numbering of the diagram. The locations of the methyl groups in *X. laevis* RNA are denoted by "m" for 2'-O-methyl groups and "M" for base methylations; see text for details. Brackets at positions 633-636 and 1281-1291 denote 2'-O-methyl groups which are not yet precisely located. One 2'-O-methyl group is at position 1353 or 1356. Pseudouridine residues so far localised (30) occur at positions 34 or 36, 105, one of positions 109-112, one of 115-120, 124, one of 136-141, and 777.

In contrast to the four helices just described, the discovery of fragments 4a and 4b (Helix 26A, Fig. 3) was a surprise. The sequence concerned would have been expected to represent one component of Helix 25 and/or Helix 19 as compared with E. coli 16S RNA (cf. Fig. 4, and ref. 20), but instead the fragments clearly form a hairpin loop, which is supported by compensating base changes in S. cerevisiae (Fig. 3). Helix 26A therefore indicates either that the eukaryotic RNA has a different structure in this region, or that there is a "switch" involved. The evidence in favour of the latter possibility will be discussed below, following a description of the revised 18S RNA secondary structure model as a whole.

Features of the X. laevis 18S model

The revised secondary structure model for the X. laevis 18S RNA is shown in Fig. 4. The model was refined from our previous proposal (9) using the fragment data just discussed, together with improved phylogenetic comparisons with the available 18S RNA sequences, namely those of S. cerevisiae (27) and rat (26), and (very recently) that of Dictyostelium discoideum (29). In addition, mutual phylogenetic comparisons with other small subunit RNA species (20) helped to establish both the 18S RNA structure and that of E. coli 16S RNA, the latest version of which (20) is also depicted for comparison in Fig. 4.

The 3'-domain of the 18S RNA model (Fig. 4B) is extremely similar to its E. coli counterpart. Helices 28-32, 35-38, 40, 43 and 45 are identical in both molecules to within at most one or two base pairs. Helix 33 (which is very variable (20)) is somewhat different in the 18S RNA, and Helices 39, 41 and 44 are a little longer. In all four of these latter features, the structure in Fig. 4 is supported by compensating base changes between S. cerevisiae, D. discoideum and X. laevis, and occasionally between rat and X. laevis. Helix 39 contains a long extra insert in D. discoideum (29). Helix 34 is essentially similar in 16S and 18S RNA, although the base-pairing is arranged slightly differently, and in Helix 42 the upper five base pairs in the 16S RNA (which can form in some but not all sequences (20)) are absent in the 18S RNA molecule.

In contrast, some parts of the 5'-domain of the 18S RNA are

still problematic. In Fig. 4A, Helices 1-3, 5, 11, 15, 18, 23, 24 and 27 are virtually identical in 16S and 18S RNA, to within one or two base pairs. Structures corresponding to the 16S RNA Helices 8, 12, 13 and 14 can also be drawn for X. laevis but are appreciably less stable in the latter. The primary sequence encompassing the elements 12 - 14 is identical in S. cerevisiae, so that no compensating base changes between different 18S RNA species can be observed in this region. In our previous model (9) we had identified Helices 11A and 11B as the 18S RNA counterparts to Helices 11 and 12 in 16S RNA, respectively, but application of the stringent homology criterion of ref. 20 indicates that the version proposed by the Strasbourg (10) and Moscow (12) groups for the region of Helices 11-15 is correct. It follows that loops 11A and 11B represent insertions in the 18S sequence, Helix 11A being experimentally confirmed (Fig. 3), and Helix 11B being supported by compensating base changes in both S. cerevisiae (9), rat and D. discoideum RNA. (Additional base pairs can be drawn between bases 271-275 and 309-313 (cf.9), but these have not yet found phylogenetic support).

This version of the structure for Helices 11-15 however causes difficulty in correlating the 18S and 16S molecules in the region of E. coli Helices 16, 17 and 4, since there is now effectively a deletion of 27 bases in the eukaryotic region corresponding to E. coli Helices 16 and 17 (10). (A recent report (31) describing an insertion of 25 bases here in eukaryotes is incorrect). The 18S RNA helix formed by bases 492-523 (Fig. 4) is supported by compensating base changes in both S. cerevisiae (9) and D. discoideum, and appears to be the counterpart to the 16S RNA Helix 16. A short helix representing a truncated version of the 16S RNA Helix 17 can be drawn in all 18S RNA species (Fig. 4), but on the other hand there is no obvious counterpart for E. coli Helix 4, although several possible short double helices can be constructed here. The consequences of the changes in this region of the 18S RNA model for "switch" structures which we have proposed (1,3) will be discussed elsewhere (R.Brimacombe, manuscript in preparation).

Helices 6, 6A and 7 are also modifications of the previous 18S RNA model (9), and the new structure proposals for X. laevis

Figure 5: Details of Helices 6, 6A (left) and 7 (right) in the *X. laevis* RNA structure (cf. Fig. 4). The loops proposed for *X. laevis* are shown, with base changes in *S. cerevisiae* (27) superimposed, as in Fig. 3. The regions of the *S. cerevisiae* sequence concerned are bases 56-101 (Helices 6, 6A), and bases 115-148, 187-220 (Helix 7).

and *S. cerevisiae* are illustrated in Fig. 5. Although these helices should still be regarded as tentative, we think that they show a closer resemblance to the corresponding *E. coli* structures, both in terms of compensating base changes and location in the primary sequence, as compared with the structures proposed by other authors (29, 31). In the same region of the molecule, the *E. coli* Helix 9 (which is not universal (20)) cannot form in 18S RNA, and the tentative Helix 10 in *E. coli* (20) appears to have been deleted, as is the case in chloroplast 16S RNA (9). Rat 18S RNA has an eleven-base insertion at position 194 (26), which can either extend Helix 7, or form a small side-loop (cf. Helix 6A, Fig. 5).

In the central region of the 18S RNA molecule, Helices 21A-21F represent the major insertion in the 18S relative to the 16S RNA. These structures are of course tentative, but it is encouraging that at least one of the loops (Helix 21C, Fig. 3) is now clearly confirmed by the fragment data. The precise nature of these extra loops could also influence the positioning of Helices 20 and 22, which are very similar to but not directly superimposable upon their *E. coli* counterparts in the model shown in Fig. 4. The putative *E. coli* Helix 19 is also in this region, but we have already noted the lack of positive evidence for this interaction (20), and it has no plausible counterpart in the 18S RNA (cf. ref. 12) if the stringent homology criterion of ref. 20 is applied.

Figure 6: A "switch" structure in *E. coli* 16S RNA. The left-hand diagram shows Helices 25 and 26 of the *E. coli* structure (cf. Fig. 4), with base changes in the sequence of *A. nidulans* mitochondrial RNA (32) superimposed. See Legend to Fig. 3 for explanation of symbols. The region of the *A. nidulans* sequence concerned comprises bases 737-818. In the right-hand diagram, Helix 25 has been opened, and re-folded into a loop corresponding to Helix 26A of the *X. laevis* structure (cf. Figs. 3 and 4).

A phylogenetically conserved "switch" structure

In 16S RNA, Helices 25 and 26 are both strongly supported by phylogenetic (20) and fragment (3) evidence. In 18S RNA Helix 26 is supported by compensating base changes between *S. cerevisiae* (9), *D. discoideum* and *X. laevis*, and a short helix similar but not identical to Helix 25 can also be drawn (9). However, as already mentioned above, the fragment data provide evidence for an entirely different helix (26A, Fig. 3) in place of Helix 25 in the 18S RNA, and a similar helix has also been proposed for *D. discoideum* (29). This raises the question as to whether a similar structure can be formed in 16S RNA. In Fig. 6, the *E. coli* Helices 25 and 26 are shown, compared with the corresponding regions of *Aspergillus nidulans* mitochondrial 15S RNA (32). The latter sequence has many base substitutions in this region, and as drawn in Fig. 6 (left) these base changes are almost all compensating. (It should be noted that the secondary structure element proposed by Köchel and Kuntzel (32) for this region is different, and does not correspond well to the

E. coli structure). Fig. 6 (right) shows clearly however that Helix 25 can be opened up, and re-folded to give a loop which corresponds to Helix 26A. Six compensating base changes can be observed in this loop with respect to the A. nidulans sequence, and the sequence A-A-G-U-C at the loop end in E. coli is homologous with the loop end of Helix 26A in X. laevis and S. cerevisiae (Fig. 3). This strongly suggests that a "switch" can indeed exist between Helices 25 and 26A. In other small subunit 16S or 15S RNA species (cf. ref. 20) the base-pairing in the latter helix is in general conserved, with the exceptions of Halobacterium volcanii (33) and S. cerevisiae mitochondrion (34), but in both these cases a hairpin loop similar to that of E. coli can be drawn in the regions corresponding to bases 860-886 (Fig. 6, right), with the sequence A-A-G at the loop end.

RNA methylation

Ribosomal RNA from eukaryotes is relatively highly methylated. X. laevis 18S RNA contains 40 methyl groups; most of these are 2'-O-ribose substituents and the remaining few are attached to bases (16). It is known for representative eukaryotes that nearly all 2'-O-methylation of ribosomal RNA sequences occurs rapidly within the precursor RNA in the nucleolus, whereas the 18S RNA base methyl groups are added later during ribosome maturation (reviewed in ref. 35). This pattern of events can reasonably also be inferred for X. laevis. Until recently the locations of the methyl groups in the eukaryotic RNA were largely unknown, and this severely constrained attempts to understand the phenomenon of ribosomal RNA methylation. Determination of the X. laevis 18S rDNA primary structure was closely coupled to locating the methyl groups on the corresponding RNA, as described in outline in ref. 7 (a detailed account is in preparation by B.E.H. Maden). Although in a few instances the locations are uncertain within limits of a few nucleotides, the great majority of the locations are precise. The methylation sites are shown in relation to the secondary structure model in Fig. 4. Several interesting points may be noted, starting at the 3'-end.

The "universal" dimethyl-A sequence (e.g. 36) occurs where expected at the top of the 3'-terminal Helix 45. m⁶A occurs in the non-helical region between Helices 44 and 45. The first

surprise concerns m^7G , since in E. coli a unique m^7G occurs in the loop of Helix 18 in the 5'-region (36). Although the sequence of this loop is highly conserved between E. coli and eukaryotes, X. laevis 18S RNA does not contain an m^7G in this position. Instead, a base with the properties of m^7G occurs in the 3'-domain in a short non-helical region between Helix 42 and the long-range interaction comprising Helix 29. The last of the base methyl groups is part of the hypermodified nucleotide $m^7\Psi$ (17). This is located on the loop of Helix 31, another very highly conserved region which also contains a methylated residue in E. coli (36). All of these base methylation sites are in the 3'-domain of the molecule. The fact that they are modified late during ribosome maturation implies that they are located near the surface of the small ribosomal subunit.

For the numerous 2'-O-methylations the picture is more complicated. These substituents occur throughout the molecule, apparently in a variety of local environments. Some are outside the main helical regions; others are in hairpin loop ends or at boundaries or imperfections in helices. However, by no means all such non-helical regions or helix imperfections are punctuated by 2'-O-methylation. One generalisation seems to be that hardly any are located internally in stable helices. Perhaps the most striking aspect of the 2'-O-methyl groups is their high abundance in the 5'-region of the molecule. This region is unmodified in E. coli, with the exception of the m^7G already mentioned. Moreover, in E. coli much of this region is relatively inaccessible to nucleases, suggesting a rather tight tertiary structure (37). Much of the region shows a very similar overall secondary structure in X. laevis, but some of the helical elements are weaker than their E. coli counterparts, as mentioned above. The 2'-O-methyl groups in the X. laevis 5'-region are distributed between these imperfect helices and the non-helical regions in a rather complex way. One attractive hypothesis is that this whole region is stabilised by tertiary structure in Xenopus as in E. coli, and that a combination of secondary and tertiary interactions in the RNA generates the distinctive ribose orientations that are recognized as substrates for 2'-O-methylation.

The following lines of experiment might provide further insight into the relationship between RNA methylation and conformation: identification of the methylation sites in ribosomal RNA from other eukaryotes, identification of the sites of the many pseudouridine residues in eukaryotic RNA, and exploration of the accessibility of the modified nucleotides to chemical or enzymatic probes. With regard to the pseudouridine residues, only a few of the forty or so of these modified nucleotides have as yet been located in *X. laevis* 18S RNA (see Legend to Fig. 4). The rest remain to be distinguished from uridine, and the eventual findings may significantly add to our view of the fine features of conformation in eukaryotic (and also by inference prokaryotic) ribosomal RNA.

ACKNOWLEDGEMENTS

The authors are grateful to Dr. H.G. Wittmann for his continued support. This work was in part supported by a grant from the Medical Research Council.

*Present address: Department of Biochemistry, University of Liverpool, PO Box 147, Liverpool L69 3BX, UK

REFERENCES

1. Brimacombe, R., Maly, P. and Zwieb, C. (1983) *Progr. Nucleic Acids Res.* 28, 1-48.
2. Ross, A. and Brimacombe, R. (1979) *Nature* 281, 271-276.
3. Glotz, C. and Brimacombe, R. (1980) *Nucleic Acids Res.* 8, 2377-2395.
4. Glotz, C., Zwieb, C., Brimacombe, R., Edwards, K. and Kössel, H. (1981) *Nucleic Acids Res.* 9, 3287-3306.
5. Salim, M. and Maden, B.E.H. (1981) *Nature*, 291, 205-208.
6. Maden, B.E.H., Forbes, J.M., Stewart, M.A. and Eason, R. (1982) *EMBO J.* 1, 597-601.
7. Maden, B.E.H. (1982) in "The Cell Nucleus, Vol. X" (H. Busch and L. Rothblum, eds.), pp 319-350, Academic Press, New York.
8. Fox, G. and Woese, C.R. (1975) *Nature* 256, 505-507.
9. Zwieb, C., Glotz, C. and Brimacombe, R. (1981) *Nucleic Acids Res.* 9, 3621-3640.
10. Stiegler, P., Carbon, P., Ebel, J.P. and Ehresmann, C. (1981) *Eur. J. Biochem.* 120, 487-495.
11. Mankin, A.S., Kopylov, A.M., Rubtsov, P.M. and Skryabin, K.G. (1981) *Dokl. Akad. Nauk. USSR*, 256, 1006-1010.
12. Mankin, A.S., Kopylov, A.M. and Bogdanov, A.A. (1981) *FEBS Lett.* 134, 11-14.
13. Veldman, G.M., Klootwijk, J., de Regt, V.C.H.F., Planta,

- R.J., Branlant, C., Krol, A. and Ebel, J.P. (1981) *Nucleic Acids Res.* 9, 6935-6952.
14. Klotwijk, J. and Planta, R.J. (1973) *Eur. J. Biochem.* 39, 325-333.
15. Maden, B.E.H. and Salim, M. (1974) *J. Mol. Biol.* 88, 133-164.
16. Khan, M.S.N., Salim, M. and Maden, B.E.H. (1978) *Biochem. J.* 169, 531-542.
17. Brand, R.C., Klotwijk, J., Planta, R.J. and Maden, B.E.H. (1978) *Biochem. J.* 169, 71-77.
18. Hughes, D.G. and Maden, B.E.H. (1978) *Biochem. J.* 171, 781-786.
19. Thomas, G., Gordon, J. and Rogg, H. (1978) *J. Biol. Chem.* 253, 1101-1105.
20. Maly, P. and Brimacombe, R. (1983) *Nucleic Acids Res.* 11, 7263-7286.
21. Maden, B.E.H. and Reeder, R. (1979) *Nucleic Acids Res.* 6, 817-831.
22. Maxam, A.M. and Gilbert, W. (1977) *Proc. Nat. Acad. Sci. USA*, 74, 560-564.
23. Ross, A. and Brimacombe, R. (1978) *Nucleic Acids Res.* 5, 241-256.
24. Donis-Keller, H., Maxam, A.M. and Gilbert, W. (1977) *Nucleic Acids Res.* 4, 2527-2538.
25. Szekely, M. and Sanger, F. (1969) *J. Mol. Biol.* 43, 607-617.
26. Torczynski, R., Bollon, A.P. and Fuke, M. (1983) *Nucleic Acids Res.* 11, 4879-4890.
27. Rubtsov, P.M., Musakhanov, M.M., Zakharyev, V.M., Krayev, A.S., Skryabin, K.G. and Bayev, A.A. (1980) *Nucleic Acids Res.* 8, 5779-5794.
28. Douthwaite, S., Christensen, A. and Garrett, R.A. (1983) *J. Mol. Biol.* 169, 249-279.
29. Olsen, G.J., McCarroll, R. and Sogin, M.L. (1983) *Nucleic Acids Res.* 11, 8037-8049.
30. Salim, M. and Maden, B.E.H. (1980) *Nucleic Acids Res.* 8, 2871-2884.
31. Chan, Y.L., Gutell, R., Noller, H.F. and Wool. I.G. (1984) *J. Biol. Chem.* 259, 224-230.
32. Köchel, H.G. and Kuntzel, H. (1981) *Nucleic Acids Res.* 9, 5689-5696.
33. Gupta, R., Lanter, J.M. and Woese, C.R. (1983) *Science*, 221, 656-659.
34. Sor, F. and Fukuhara, H. (1980) *C.R. Acad. Sci. Paris, D.* 291, 933-936.
35. Maden, B.E.H. (1979) in "Transmethylation" (E. Usdin, R.T. Borchardt and C.R. Crevelling, eds.) pp 381-388, Elsevier/-North Holland Press, Amsterdam.
36. Carbon, P., Ehresmann, C., Ehresmann, B. and Ebel, J.P. (1979) *Eur. J. Biochem.* 100, 399-410.
37. Ehresmann, C., Stiegler, P., Carbon, P., Ungewickell, E. and Garrett, R.A. (1980) *Eur. J. Biochem.* 103, 439-446.